Social is bad for search, and search is bad for social

John Nagle

SiteTruth

www.sitetruth.com

Abstract

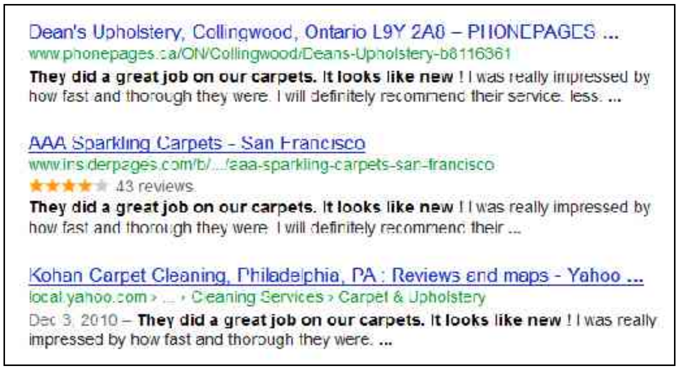
In the last two years, the concept of "social" inputs to web search has been heavily promoted. We show that social inputs to search encourage spamming to the point that search quality degrades. These attempts to pollute search are filling the "social" world with junk. An entire ecosystem has come into being to assist with search engine social spamming. Fighting this ecosystem is possible but not easy.

A tour of the world of social search spam

Our 2010 paper, "Places spam, the new front in the spam wars" was devoted to the spamming of "Google Places". Back in October 2010, Google started merging "Places" results into web search results. Spamming Google Places was known to be easy, but until last October, few people bothered, because spamming the search engine for Google Maps wasn't worth much. After the merger into web results, search engine optimization (SEO)-generated places spam via social inputs went mainstream.

The success of spam attacks on Google Places emboldened the search engine optimization industry. Previously, fear of reprisal from Google had restrained SEOs from using more aggressive "black hat" techniques to improve search positioning. After it became clear that spamming Google Places was cheap, easy, and not vulnerable to reprisals, the floodgates opened. SEO practitioners realized that aggressive techniques were now necessary to survive. That brought the industry to where it is today.

Fake reviews



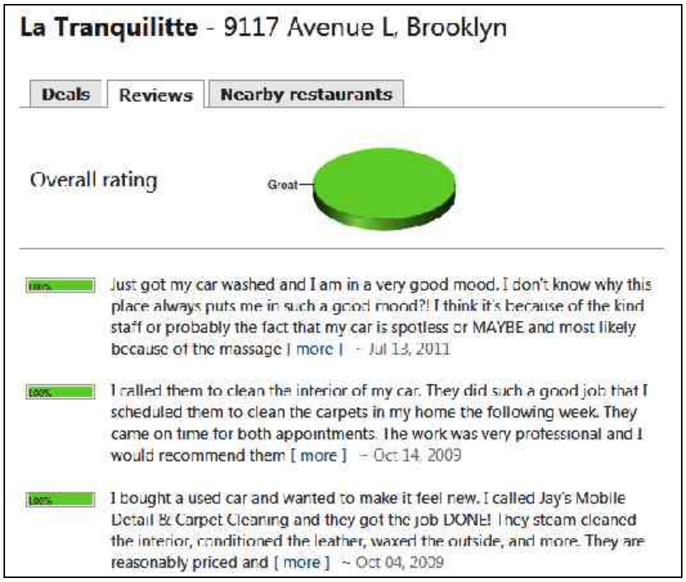
Dean's Upholstery, Collingwood, Ontario L9Y 2A8 – PHONEPAGES ...
www.phonepages.ca/ON/Collingwood/Deans-Upholstery-b8116361
They did a great job on our carpets. It looks like new ! I was really impressed by how fast and thorough they were. I will definitely recommend their service. less. ...

AAA Sparkling Carpets - San Francisco
www.insiderpages.com/b/.../aaa-sparkling-carpets-san-francisco
★★★★☆ 43 reviews
They did a great job on our carpets. It looks like new ! I was really impressed by how fast and thorough they were. I will definitely recommend their ...

Kohan Carpet Cleaning, Philadelphia, PA : Reviews and maps - Yahoo ...
local.yahoo.com › ... › Cleaning Services › Carpet & Upholstery
Dec 3, 2010 – **They did a great job on our carpets. It looks like new** ! I was really impressed by how fast and thorough they were. ...

Fake reviews are common. Here, a search for the phrase ***"They did a great job on our carpets. It looks like new"*** brings up the same review text for different carpet cleaning companies in different cities. This is low-end search engine optimization in action.

An entire industry has sprung up to generate such fake reviews. Some SEO practitioners are better at it than others. The Wikizip.com entry for "La Tranquilitte", a restaurant in Brooklyn, NY, demonstrates unusual SEO incompetence; reviews of a car wash have been applied to a restaurant's entry.



La Tranquilitte - 9117 Avenue L, Brooklyn

100% Just got my car washed and I am in a very good mood. I don't know why this place always puts me in such a good mood?! I think it's because of the kind staff or probably the fact that my car is spotless or MAYBE and most likely because of the massage | more | - Jul 13, 2011

100% I called them to clean the interior of my car. They did such a good job that I scheduled them to clean the carpets in my home the following week. They came on time for both appointments. The work was very professional and I would recommend them [more] - Oct 14, 2009

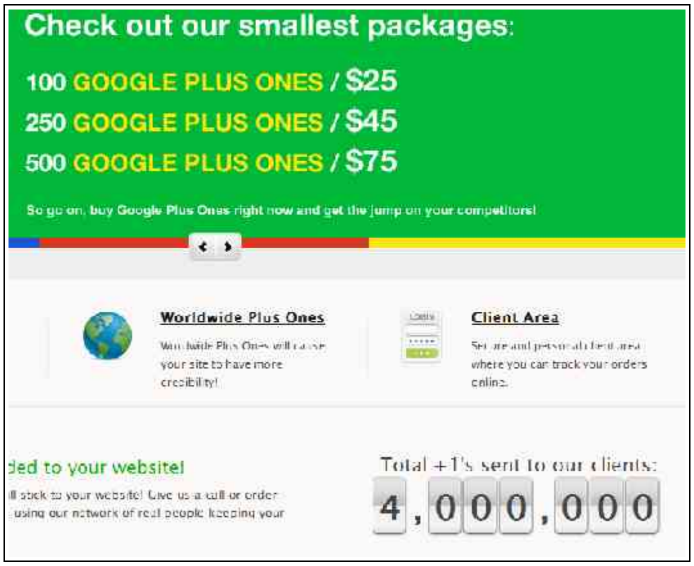
100% I bought a used car and wanted to make it feel new. I called Jay's Mobile Detail & Carpet Cleaning and they got the job DONE! They steam cleaned the interior, conditioned the leather, waxed the outside, and more. They are reasonably priced and [more] ~ Oct 04, 2009

Such reviews are typically machine-generated by scraping reviews from other sites and repurposing them. Most spam of this type, though, at least copies reviews from a similar business.

Those fake reviews aren't intended to be read by people. They're just there to be counted by search engines. So such seemingly nonsensical reviews have value.

The same is true of "likes", and "+1″s, which result in the creation of bogus social accounts for spamming purposes. Too much spam can kill a social network. That's part of what happened to Myspace, Craigslist, almost every online dating site, and now, it is hitting Facebook and Google.

Fake "Likes" and "+1"s.



Check out our smallest packages:

100 GOOGLE PLUS ONES / $25
250 GOOGLE PLUS ONES / $45
500 GOOGLE PLUS ONES / $75

So go on, buy Google Plus Ones right now and get the jump on your competitors!

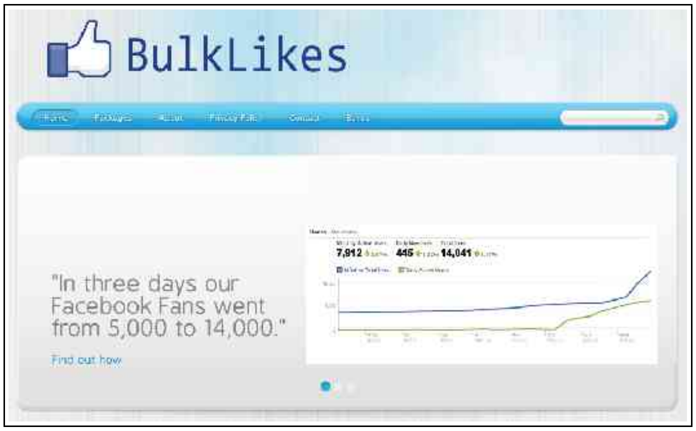
Worldwide Plus Ones
Worldwide Plus Ones will cause your site to have more credibility!

Client Area
Secure and personalized client area where you can track your orders online.

ded to your website!
ll stick to your website! Give us a call or order
using our network of real people keeping your

Total +1's sent to our clients:
4,000,000

The ad to the left is from "googleplus1supply.com", and shows the going rates for a Google "+1" boost. The fake "+1" industry is highly competitive. Other vendors include "plus1sem.com" ("Buy 2000 Google Plus Ones and SKYROCKET your rankings") "buyplus1fans.com" ("Our service helps boost your Google +1 presence which will convert into higher rankings equaling more customers!"), and "buyrealplusone.com" ("Crush your competition!").



BulkLikes

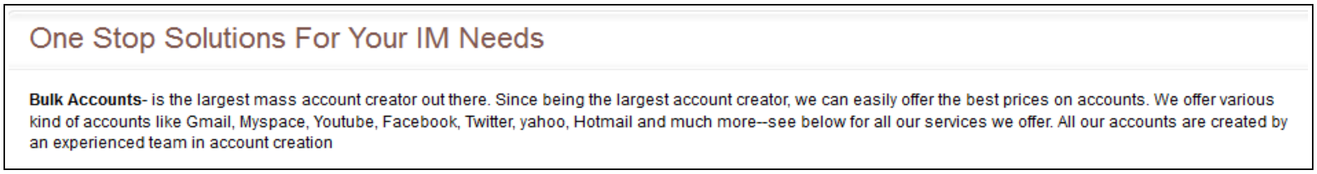
"In three days our Facebook Fans went from 5,000 to 14,000."

Find out how

Facebook's "likes" are also available in bulk. "bulklikes.com" offers 500 Facebook "fans" for $260. Competitors include "premiumfans.net" ("Buy Facebook Fans with us today and watch your popularity boom.") and "buyfacebulkfans.com" (1-855-BUY-FANS")

Fake users

The fake “like” and “+1” business requires the creation of fake accounts on social networks. For that, too, there is an industry.

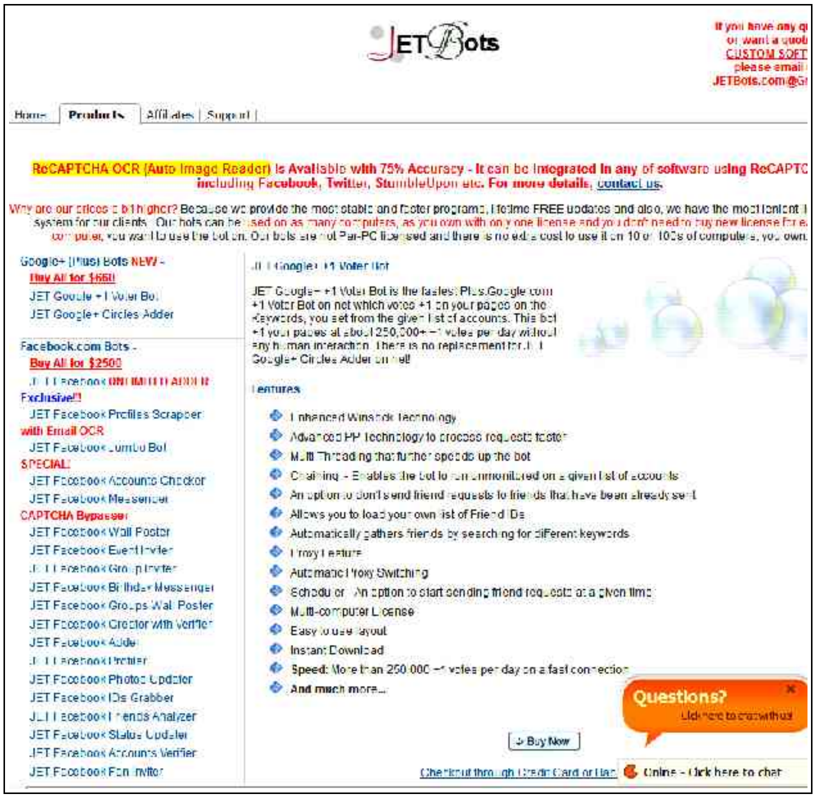


One Stop Solutions For Your IM Needs

Bulk Accounts- is the largest mass account creator out there. Since being the largest account creator, we can easily offer the best prices on accounts. We offer various kind of accounts like Gmail, Myspace, Youtube, Facebook, Twitter, yahoo, Hotmail and much more--see below for all our services we offer. All our accounts are created by an experienced team in account creation

The ad above is from “bulkaccounts.com”, based in India and Australia. They offer fake accounts on Gmail, Myspace, Facebook, Youtube, Twitter, Yahoo, and Hotmail. Theirs is a semi-automated service using low-wage labor.

Unlike the “fake +1” services, fake account services are marketed not to end users or advertisers, but to those in the SEO industry. These tools and services are not widely publicized, but are mentioned on “black hat” forums. The businesses behind them tend to guard their anonymity. Few list a business address.



ReCAPTCHA OCR (Auto Image Reader) is Available with 75% Accuracy - It can be Integrated in any of software using ReCAPTC including Facebook, Twitter, StumbleUpon etc. For more details, contact us.

Why are our prices a bit higher? Because we provide the most stable and faster programs, lifetime FREE updates and also, we have the most lenient l system for our clients. Our bots can be used on as many computers, as you own with only one license and you don't need to buy new license for e computer, you want to use the bot on. Our bots are not Per-PC licensed and there is no extra cost to use it on 10 or 100s of computers, you own.

Google+ (Plus) Bots NEW -
Buy All for $650
JET Google +1 Voter Bot
JET Google+ Circles Adder

Facebook.com Bots -
Buy All for $2500
JET Facebook UNLIMITED ADDER
Exclusive!!
JET Facebook Profiles Scrapper
with Email OCR
JET Facebook Jumbo Bot
SPECIAL
JET Facebook Accounts Checker
JET Facebook Messenger
CAPTCHA Bypasser
JET Facebook Wall Poster
JET Facebook Event Inviter
JET Facebook Group Inviter
JET Facebook Birthday Messenger
JET Facebook Groups Wall Poster
JET Facebook Creator with Verifier
JET Facebook Adder
JET Facebook Profiler
JET Facebook Photos Updater
JET Facebook IDs Grabber
JET Facebook Friends Analyzer
JET Facebook Status Updater
JET Facebook Accounts Verifier
JET Facebook Fan Inviter

JET Google+ +1 Voter Bot

JET Google+ +1 Voter Bot is the fastest Plus.Google.com +1 Voter Bot on net which votes +1 on your pages on the Keywords, you set from the given list of accounts. This bot +1 your pages at about 250,000+ +1 votes per day without any human interaction. There is no replacement for JET Google+ Circles Adder on net!

Features

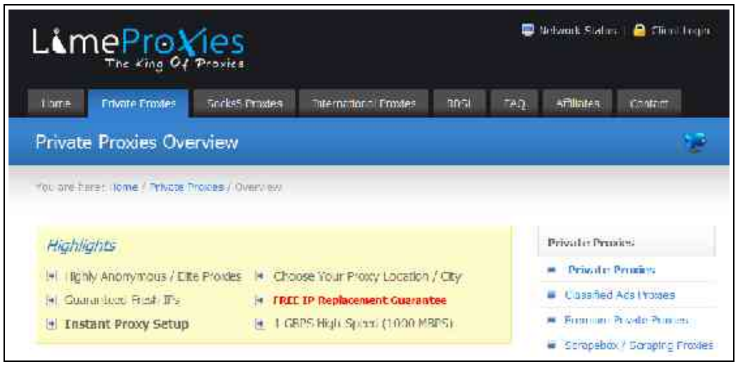
- Enhanced Winsock Technology
- Advanced PP Technology to process requests faster
- Multi Threading that further speeds up the bot
- Chaining - Enables the bot to run unmonitored on a given list of accounts
- An option to don't send friend requests to friends that have been already sent
- Allows you to load your own list of Friend IDs
- Automatically gathers friends by searching for diFerent keywords
- Proxy Feature
- Automatic Proxy Switching
- Scheduler - An option to start sending friend requests at a given time
- Multi-computer License
- Easy to use layout
- Instant Download
- **Speed:** More than 250,000 +1 votes per day on a fast connection
- **And much more...**

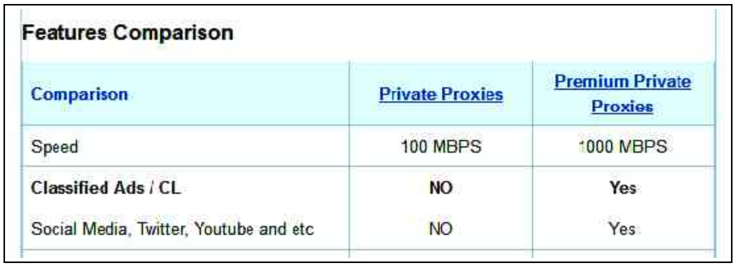
JetBots, advertised at left, can create fake Facebook and Google accounts. This program creates accounts, bypassing CAPTCHAs using a combination of advanced optical character recognition technology and outsourcing to low-wage countries. Once the account has been created, it adds plausible profile information. Then the spamming begins.

Facebook spamming tools have been available for some time. It took a few months for the web spam industry to crack Google+. JetBots now advertises tools for spamming Google+ as well as Facebook.

With such power tools, spam can be generated in much greater volume than with the manual, outsourced services. JetBots advertises “250,000 +1 votes per day on a fast connection”.

Fake Internet Protocol (IP) addresses





Features Comparison

Comparison	Private Proxies	Premium Private Proxies
Speed	100 MBPS	1000 MBPS
Classified Ads / CL	NO	Yes
Social Media, Twitter, Youtube and etc	NO	Yes

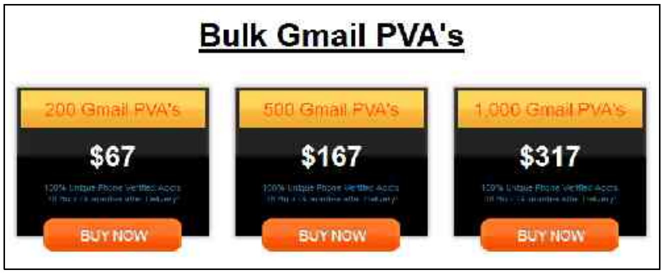
The creation of fake accounts can be detected by services which log the Internet adddress (IP address) from which the request comes. Thus, the bulk creation of fake accounts, fake reviews, and fake postings requires fake IP addresses. Such IP addresses can be rented from “proxy services”, such as “LimeProxies.com”.

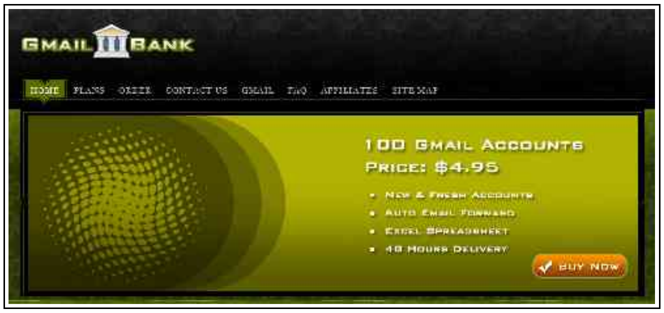
This vendor is clear about the purposes of their service. “Premium Private Proxies” are for use with “Craigslist, social media, Twitter, Youtube and etc.” Other vendors include “ezproxies.com” and “getfoxyproxy.org”. All these vendors are Google advertisers.

The proxy business is partly legitimate and partly a front for organized crime. Some “proxies” are computers which have been broken into remotely and taken over by a “botnet”. Much spam is sent out through such compromised machines, and they are also used for credit card fraud. To purchase botnet proxies, one has to look in less reputable places, such as the forums of “black hat SEO” sites. These are not difficult to find. We will not provide details here.

Fake e-mail accounts

Most social networking services require a unique e-mail account to sign up. Social spam SEOs must thus acquire large numbers of fake e-mail accounts. Until 2010, the most popular tool for creating bulk accounts was “Jiffy Gmail Creator”, but due to changes at Gmail, it no longer works.


Bulk Gmail PVA's
200 Gmail PVA's
$67
100% Unique Phone Verified Accts
BUY NOW
500 Gmail PVA's
$167
100% Unique Phone Verified Accts
BUY NOW
1,000 Gmail PVA's
$317
100% Unique Phone Verified Accts
BUY NOW

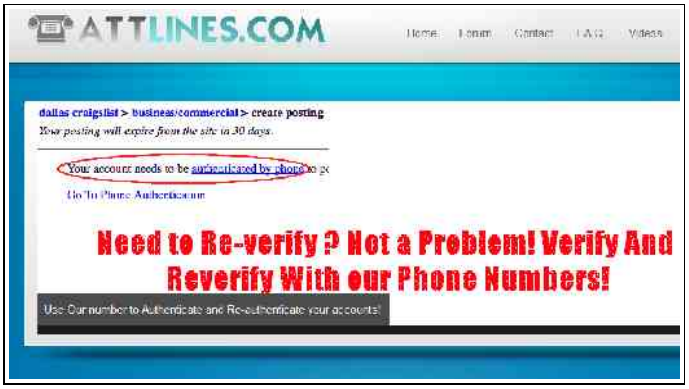


Other services have taken up the slack. This ad, from “xgcmedia.com”, offers 1000 fake phone-verified (“PVA”) Gmail accounts for $317. Gmail seems to be the preferred provider for this purpose. Hotmail and Yahoo accounts are also available in bulk, but they are considered much less valuable and are priced lower.

Gmail Bank offers 100 Gmail accounts for $4.95, but these are not “phone verified”, and are less valuable.

Fake phone verified accounts require fake phone numbers, which we will cover next.

Fake phone numbers



dallas craigslist > business/commercial > create posting

Your posting will expire from the site in 30 days.

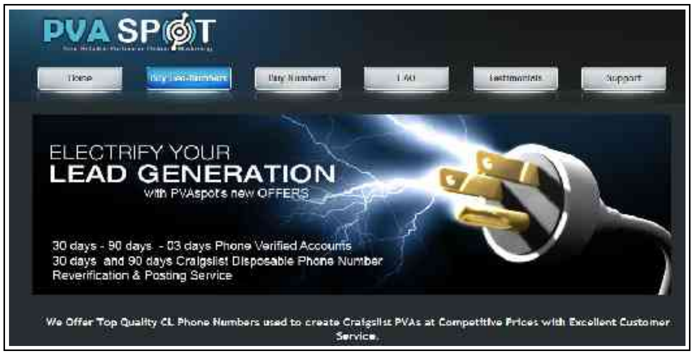
Your account needs to be authenticated by phone to go

Go To Phone Authentication

Need to Re-verify ? Not a Problem! Verify And Reverify With our Phone Numbers!

Use Our number to Authenticate and Re-authenticate your accounts!

Social networks have tried to fight fake accounts. Craigslist has tried CAPTCHAs, email verification, and even verification by telephone. Requiring a unique phone number for each user (a "Phone Verified Account", or PVA) created demand for fake phone numbers. This service, "attlines.com", creates fake phone numbers to support fake online identities. Each phone number is valid for only one month, and is allowed only 20 minutes of talk time.



ELECTRIFY YOUR LEAD GENERATION with PVAspots new OFFERS

30 days - 90 days - 03 days Phone Verified Accounts
30 days and 90 days Craigslist Disposable Phone Number
Reverification & Posting Service

We Offer Top Quality CL Phone Numbers used to create Craigslist PVAs at Competitive Prices with Excellent Customer Service.

A competing service, "pvaspot.com", is quite clear about their role in attacking Craigslist: "Top Quality CL Phone Numbers used to create Craigslist PVAs".

Again, these are services sold to SEOs, not end users.

This is just an overview of the situation. There's also fake business location spamming (covered in our previous paper), Twitter spamming, and blog and forum spamming. If a social networking system can be profitably spammed, it is being spammed.

Traditional search spamming, such as link farms and spam blogs, continues to be popular. However, it involves ongoing expense. Link farms and spam blogs require hosting. Social spam is hosted for free by the social networks, so costs are lower and there's little risk of a site being blocked.

The social spam ecosystem

The social spam ecosystem has several levels, of decreasing legitimacy. At the top are the SEO firms which offer to promote businesses. These operators usually, but not always, have a business identity a business address, can be reached by phone or mail, and accept normal forms of business payment.

These social spam services don't hide. SEO firms which offer to "enhance" search ranking through social signals operate quite openly. This is a change from the previous generation of search engine spamming, where there was real fear that Google would detect a link farm and apply a penalty which would make a business disappear from search results. Link farm operators kept a low profile. Social spam is not like that. It is currently low-risk, more businesses are willing to embrace it, and it appears to be a growth industry.

At the next level down are the companies which generate fake reviews, Facebook "likes", and Google

"+1s". While advertisers can deal with these companies directly, they are most often used by SEO firms to do the dirty work. At this level, there is usually no business address; contact is on-line only, and payment methods begin to become nonstandard.

Further down are the proxy services, fake e-mail account and fake phone number businesses. Contact and payment are typically anonymous, but the providers usually have web sites.

The botnet operators are criminal enterprises. They are moderately difficult to find, have no visible public presence, and communicate anonymously. Payment is usually through some marginally legal method. Botnet operators are pursued by law enforcement, with occasional success.

This hierarchy insulates the SEO firms at the top from the criminal activity at the bottom.

What this means for search engines

Every attempt by a major search engine to use social signals has been heavily spammed. Google Places was hit hard starting in October 2010, when Places results were mixed in with web search results. It happened fast - within two months, Google Places was choked with spam, with both phony locations and phony reviews. This was so bad that the mainstream press picked up on it, and Google had to de-emphasize "places" results. A year later Google is still being hammered in the press. The New York Times and Fox News both say Google has a problem. (When both of those sources say you have a problem, you have a problem.)

On the social side, all this spam activity has jammed social sites with junk intended for automatic reading by search engine spiders. The users of social sites see this junk and use the social service less. From a social network's point of view, all this spam activity generates cost, annoys users, and generates no revenue.

Yet search engines want to use social signals. Not because they improve search quality, but because they increase user engagement. When a search engine produces the user's desired search result on the first try, and the user immediately clicks on that result, the search engine makes no money. Search users use search sites frequently but spend little time there. Social features cause users to spend more time on search sites, exposing them to more advertising. Social signals are thus a marketing tool. They should be viewed as such.

There are useful social signals for search, but they come from systems that see transactions and know who bought something, such as Amazon, eBay, and Visa International. Even those can be spammed; spammers can buy an old eBay account, change the name, and inherit the old reputation.

Guidelines for reliable search

- **Social signals should not be used for objectively verifiable information.** The name and physical location of a business is not a matter of opinion. The line of business of a business is not a matter of opinion. The size of a business is not a matter of opinion. Social signals should not be used to establish hard data. It's tempting to try to "crowdsource" such things as a cost-saving measure, but search engines which have tried that have been inundated by entries for fake business locations. There are reliable data sources for basic business information, and those should be used instead.
- **Reviews are only meaningful if from real customers.** Counting anonymous reviews is asking

for spam. Where transaction data is available which identifies the customer and the item purchased, as with eBay, Amazon, or Visa International, reviews are known to be from real customers. Search engines may choose to display reviews to increase social engagement, but reviews from anonymous users should not affect search rankings.

- **Fake users cannot be reliably distinguished from real users.** Some social sites have tried hard to distinguish fake users from real ones, with limited success. As long as fake users can be created at low cost, this is futile. See “Inside Craigslist's Increasingly Complicated Battle Against Spammers”. Craigslist lost that battle.
- **Social signals should be applied to search ranking only when from “friends”, and only to personalized search.** “Friends” should be interpreted narrowly, as people the user actually knows and who know them. “Fans” are not friends. “Followers” are not friends. This makes social spamming via phony accounts far less useful, because its reach will be very limited.

Conclusion

Social signals are easy and cheap to spam. Over the last few years, social spammers have developed substantial infrastructure and a sizable number of commercial enterprises to generate fake data and feed it to search engines. Search engines can no longer trust social signals. Strong defensive measures are required to resist social spam.